Multivariate Statistical Analysis

多元统计分析

2026 年 4 月 9 日

已学知识点 (Recap)

第 4 章 多元分布

4.1 分布函数与密度函数

- 累积分布函数 (cdf) 的定义为 $F(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} < \boldsymbol{x})$.
- 如果概率密度函数 (pdf) 存在，则 $F(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$.
- 概率密度函数 (pdf) 满足 $\int_{-\infty}^{\infty} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$.
- 将随机向量 $\boldsymbol{X}$ 分块为 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$，设 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 的联合 (即 $\boldsymbol{X}$ 的) 分布函数为 F，则 $\boldsymbol{X}_1$ 的边际分布函数为 $F_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \mathbb{P}(\boldsymbol{X}_1 < \boldsymbol{x}_1)$，$\boldsymbol{X}_1$ 的边际概率密度函数显然为 $f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \int_{-\infty}^{\infty} f(\boldsymbol{x}_1, \boldsymbol{x}_2)\,\mathrm{d}\boldsymbol{x}_2$. 对联合概率密度函数求导亦可得到相同的边际概率密度函数.
- 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 条件概率密度函数为 $f\left(\boldsymbol{x}_2 \middle| \boldsymbol{x}_1\right) = \dfrac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}$.

已学知识点 (Recap)

第 4 章 多元分布

4.1 分布函数与密度函数

- 两个随机向量 X_1 与 X_2 独立，当且仅当 $f(\boldsymbol{x}_1, \boldsymbol{x}_2) = f_{X_1}(\boldsymbol{x}_1) \cdot f_{X_2}(\boldsymbol{x}_2)$，也等价于 $f\left(\boldsymbol{x}_2 \middle| \boldsymbol{x}_1\right) = f_{X_2}(\boldsymbol{x}_2)$.
- 不同的联合概率密度函数可以有相同的边际概率密度函数.
- **连接** (copula) 是将边际分布函数联系起来以形成联合分布函数的一个函数.

已学知识点 (Recap)

第 4 章 多元分布

4.2 矩与特征函数

- 随机向量 $\boldsymbol{X}$ 的期望为 $\boldsymbol{\mu}=\int \boldsymbol{x} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$，协方差矩阵为 $\Sigma=\mathrm{Var}(\boldsymbol{X})=\mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\right]$，将其记为 $\boldsymbol{X}\sim(\boldsymbol{\mu},\Sigma)$.
- 求期望是线性运算，即 $\mathbb{E}\left(\alpha\boldsymbol{X}+\beta\boldsymbol{Y}\right)=\alpha\mathbb{E}(\boldsymbol{X})+\beta\mathbb{E}(\boldsymbol{Y})$. 如果 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立，则 $\mathbb{E}\left(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\right)=\mathbb{E}(\boldsymbol{X})\,\mathbb{E}\left(\boldsymbol{Y}^{\mathrm{T}}\right)$.
- 两个随机向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的协方差矩阵为
$$\Sigma_{\boldsymbol{XY}}=\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})=\mathbb{E}\left[(\boldsymbol{X}-\mathbb{E}(\boldsymbol{X}))(\boldsymbol{Y}-\mathbb{E}(\boldsymbol{Y}))^{\mathrm{T}}\right]=\mathbb{E}\left(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\right)-\mathbb{E}(\boldsymbol{X})\,\mathbb{E}\left(\boldsymbol{Y}^{\mathrm{T}}\right).$$
如果 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 独立，则 $\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})=0$.
- 随机向量 $\boldsymbol{X}$ 的特征函数 (cf) 为 $\varphi_{\boldsymbol{X}}(\boldsymbol{t})=\mathbb{E}\left(\mathrm{e}^{\mathrm{i}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}}\right)$.
- p 维随机向量 $\boldsymbol{X}$ 的分布完全由形如 $\boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}$ 的所有一维分布确定，其中 $\boldsymbol{t}\in\mathbb{R}^{p}$ (Cramer-Wold 定理).
- 条件期望 $\mathbb{E}\left(X_2\,\middle|\,X_1\right)$ 是用 X_1 的一个函数对 X_2 在均方误差 (MSE) 意义下的最佳近似.

已学知识点 (Recap)

第 4 章 多元分布

4.3 变换

- 如果 $\boldsymbol{X}$ 的密度函数为 $f_X(\boldsymbol{x})$，设 $\boldsymbol{X}=u(\boldsymbol{Y})$，则变换后的随机向量 $\boldsymbol{Y}$ 的概率密度函数为 $f_Y(\boldsymbol{y})=\operatorname{abs}\left(|\mathscr{J}|\right)\cdot f_X\left[u(\boldsymbol{y})\right]$，其中 $\mathscr{J}$ 表示变换的 Jacobian 矩阵 $\mathscr{J}=\left(\frac{\partial u(y_i)}{\partial y_j}\right)$.
- 对于线性关系 $\boldsymbol{Y}=\mathscr{A}\boldsymbol{X}+\boldsymbol{b}$ 的情形，$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的概率密度函数之间的关系为
$$f_Y(\boldsymbol{y})=\operatorname{abs}\left(|\mathscr{A}|^{-1}\right)\cdot f_X\left\{\mathscr{A}^{-1}(\boldsymbol{y}-\boldsymbol{b})\right\}.$$

已学知识点 (Recap)

第 4 章 多元分布

4.4 多元正态分布

- p 维正态分布 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ 的概率密度函数为
$$f(\boldsymbol{x}) = \left|2\pi\Sigma\right|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$
多元正态分布的密度函数的等高线是椭球面，其半轴的长度与 $\sqrt{\lambda_i}$ 成比例，其中 λ_i 表示 Σ 的特征值 $(i = 1, 2, \dots, p)$.
- 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，经 Mahalanobis 变换有 $\boldsymbol{Y} = \Sigma^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}) \sim N_p\left(\boldsymbol{0}, \mathscr{I}_p\right)$. 反之，我们可以由 $\boldsymbol{Y} \sim N_p\left(\boldsymbol{0}, \mathscr{I}_p\right)$ 经过变换 $\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Y} + \boldsymbol{\mu}$ 得到 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$.
- 如果协方差矩阵 Σ 是奇异矩阵 (即，$\operatorname{rank}(\Sigma) < p$)，则它定义了一个奇异正态分布.
- 奇异正态分布的密度函数为 $\dfrac{(2\pi)^{-k/2}}{\left(\lambda_1\lambda_2\cdots\lambda_k\right)^{1/2}} \exp\left\{-\dfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$.

已学知识点 (Recap)

第 4 章 多元分布

4.5 抽样分布与中心极限定理

- 设 $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ 是独立同分布的随机向量，且 $\boldsymbol{X}_i \sim N_p(\boldsymbol{\mu}, \Sigma)$，则 $\bar{\boldsymbol{x}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\Sigma\right)$.
- 设 $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ 是独立同分布的随机向量，且 $\boldsymbol{X}_i \sim (\boldsymbol{\mu}, \Sigma)$，则 $\sqrt{n}\,(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$ 的分布渐进服从 $N(\boldsymbol{0}, \Sigma)$ (中心极限定理).
- 设 $X_1, X_2, \ldots, X_n$ 是独立同分布的随机变量，且 $X_i \sim (\mu, \sigma)$，则由中心极限定理可得 μ 的渐进置信区间为：$\bar{x} \mp \frac{\hat{\sigma}}{\sqrt{n}} u_{\frac{\alpha}{2}}$.
- 如果 $\boldsymbol{t}$ 是一个渐进服从正态分布的统计量，即，$\sqrt{n}\,(\boldsymbol{t} - \boldsymbol{\mu}) \xrightarrow{\mathscr{L}} N_p(\boldsymbol{\mu}, \Sigma)$，则该结论对函数 $f(\boldsymbol{t})$ 也成立，即，$\sqrt{n}\left\{f(\boldsymbol{t}) - f(\boldsymbol{\mu})\right\}$ 也渐进服从正态分布.

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 厚尾：尾部面积相比同均值 μ 、同方差 σ^2 的正态分布有更高的概率. 厚尾分布的峰度大于3.
- 广义双曲分布 (一维)

$$f_{\mathrm{GH}}\left(x;\ \lambda,\alpha,\beta,\delta,\mu\right)=\frac{\left(\dfrac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\ K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\cdot\frac{K_{\lambda-1/2}\left[\alpha\sqrt{\delta^2+(x-\mu)^2}\right]}{\left(\dfrac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{1/2-\lambda}}\cdot e^{\beta(x-\mu)}$$

$$K_\lambda(x)=\frac{1}{2}\int_0^\infty y^{\lambda-1}e^{-\frac{x}{2}\left(y+y^{-1}\right)}dy\ ,$$

修正的指数为 λ 第三类贝塞尔 (Bessel) 函数

参数的取值范围：
$$\begin{cases}\mu\in\mathbb{R} & \\ \delta\geq 0\ ,\quad |\beta|<\alpha\ , & \text{if } \lambda>0\\ \delta>0\ ,\quad |\beta|<\alpha\ , & \text{if } \lambda=0\\ \delta\geq 0\ ,\quad |\beta|\leq\alpha\ , & \text{if } \lambda<0\end{cases}$$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 厚尾：尾部面积相比同均值 μ 、同方差 σ^2 的正态分布有更高的概率. 厚尾分布的峰度大于3.
- 广义双曲分布 (一维)

$$f_{\mathrm{GH}}\left(x;\ \lambda,\ \alpha,\ \beta,\ \delta,\ \mu\right)=\frac{\left(\dfrac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\ K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\cdot\frac{K_{\lambda-1/2}\left[\alpha\sqrt{\delta^2+(x-\mu)^2}\right]}{\left(\dfrac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{1/2-\lambda}}\cdot \mathrm{e}^{\beta(x-\mu)}$$

$$\mathrm{E}(X)=\mu+\frac{\delta\,\beta}{\sqrt{\alpha^2-\beta^2}}\cdot\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$$

$$\mathrm{Var}(X)=\delta^2\left\{\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\delta\sqrt{\alpha^2-\beta^2}\ K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}+\frac{\beta^2}{\alpha^2-\beta^2}\left[\frac{K_{\lambda+2}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}-\left(\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right)^2\right]\right\}$$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 厚尾：尾部面积相比同均值 μ 、同方差 σ^2 的正态分布有更高的概率. 厚尾分布的峰度大于3.
- 广义双曲分布 (一维)

$$f_{\mathrm{GH}}\left(x;\ \lambda,\ \alpha,\ \beta,\ \delta,\ \mu\right)=\frac{\left(\dfrac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\ K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\cdot\frac{K_{\lambda-1/2}\left[\alpha\sqrt{\delta^2+(x-\mu)^2}\right]}{\left(\dfrac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{1/2-\lambda}}\cdot \mathrm{e}^{\beta(x-\mu)}$$

当 $\lambda=1$ 时，得到双曲 (HYP) 分布

$$f_{\mathrm{HYP}}\left(x;\ \alpha,\ \beta,\ \delta,\ \mu\right)=\frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\ \mathrm{e}^{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)},\quad x,\ \mu\in\mathbb{R},\ \delta\geq 0,\ \left|\beta\right|<\alpha$$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 厚尾：尾部面积相比同均值 μ 、同方差 σ^2 的正态分布有更高的概率. 厚尾分布的峰度大于3.
- 广义双曲分布 (一维)

$$f_{\mathrm{GH}}\left(x;\ \lambda,\,\alpha,\,\beta,\,\delta,\,\mu\right)=\frac{\left(\dfrac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\ K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\cdot\frac{K_{\lambda-1/2}\left[\alpha\sqrt{\delta^2+(x-\mu)^2}\right]}{\left(\dfrac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{1/2-\lambda}}\cdot\mathrm{e}^{\beta(x-\mu)}$$

当 $\lambda=-\dfrac{1}{2}$ 时，得到正态逆高斯 (NIG) 分布

$$f_{\mathrm{NIG}}\left(x;\ \alpha,\,\beta,\,\delta,\,\mu\right)=\frac{\alpha\delta}{\pi}\,\frac{K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\sqrt{\delta^2+(x-\mu)^2}}\ \mathrm{e}^{\delta\sqrt{\alpha^2-\beta^2}\,+\,\beta(x-\mu)}$$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 学生 t 分布：概率密度函数为

$$f_t(x;\ n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\ \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty$$

自由度

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \mathrm{e}^{-x} dx$$

均值 $\mu = 0$

方差 $\sigma^2 = \dfrac{n}{n-2}$

斜度 $= 0$

峰度 $= 3 + \dfrac{6}{n-4}$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- Laplace 分布：均值为 μ，尺度参数为 θ 的 Laplace 分布的概率密度函数为

$$f_{\text{Laplace}}(x;\ \mu,\theta) = \frac{1}{2\theta}\mathrm{e}^{-\frac{|x-\mu|}{\theta}},\quad -\infty < x < \infty$$

$$\mathrm{sign}(x-\mu) = \begin{cases} 1, & x > \mu \\ 0, & x = \mu \\ -1, & x < \mu \end{cases}$$

分布函数为

$$F_{\text{Laplace}}(x;\ \mu,\theta) = \frac{1}{2}\left[1 + \mathrm{sign}(x-\mu)\left(1 - \mathrm{e}^{-\frac{|x-\mu|}{\theta}}\right)\right]$$

$$\text{均值 } \mu = \mu$$
$$\text{方差 } \sigma^2 = 2\theta^2$$
$$\text{斜度} = 0$$
$$\text{峰度} = 6$$

标准 Laplace 分布：$\mu = 0$，$\theta = 1$

$$f(x) = \frac{\mathrm{e}^{-|x|}}{2}$$

$$F(x) = \begin{cases} \dfrac{\mathrm{e}^{x}}{2}, & x < 0 \\ 1 - \dfrac{\mathrm{e}^{-x}}{2}, & x \geq 0 \end{cases}$$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- Cauchy 分布：位置参数 m，尺度参数 s 的 Cauchy 分布的概率密度函数和分布函数分别为

$$
\begin{aligned}
f_{\text{Cauchy}}(x;\ m,\ s) &= \frac{1}{s\pi}\cdot\frac{1}{1+\left(\frac{x-m}{s}\right)^2} \\
F_{\text{Cauchy}}(x;\ m,\ s) &= \frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{x-m}{s}\right)
\end{aligned}
\quad ,\quad -\infty < x < \infty
$$

- 标准 Cauchy 分布：$m = 0,\ \ s = 1$

$$
\begin{aligned}
f_{\text{Cauchy}}(x) &= \frac{1}{\pi\left(1+x^2\right)} \\
F_{\text{Cauchy}}(x) &= \frac{1}{2}+\frac{\arctan x}{\pi}
\end{aligned}
\quad ,\quad -\infty < x < \infty
$$

- Cauchy 分布的均值、方差、斜度、峰度不存在.
- Cauchy 分布的众数 (mode) 和中位数存在，都等于位置参数 m.

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 混合模型：由 L 个分布构成的**混合分布** (mixture distribution) 的概率密度函数可表示为

权重

$$f(x) = \sum_{l=1}^{L} \boxed{w_l} \cdot p_l(x)$$

约束条件：$0 \leq w_l \leq 1$， $\sum_{l=1}^{L} w_l = 1$， $\int p_l(x) dx = 1$

均值 $\mu = \sum_{l=1}^{L} w_l \mu_l$

方差 $\sigma^2 = \sum_{l=1}^{L} w_l \left[\sigma_l^2 + (\mu_l - \mu)^2 \right]$

斜度 $= \sum_{l=1}^{L} w_l \left[\left(\frac{\sigma_l}{\sigma} \right)^3 SK_l + \frac{3\sigma_l^2 (\mu_l - \mu)}{\sigma^3} + \left(\frac{\mu_l - \mu}{\sigma} \right)^3 \right]$

峰度 $= \sum_{l=1}^{L} w_l \left[\left(\frac{\sigma_l}{\sigma} \right)^4 K_l + \frac{6 (\mu_l - \mu)^2 \sigma_l^2}{\sigma^4} + \frac{4 (\mu_l - \mu) \sigma_l^3}{\sigma^4} SK_l + \left(\frac{\mu_l - \mu}{\sigma} \right)^4 \right]$

只要混合分布中密度函数的数量足够多，并且模型的参数选择正确的话，混合模型就能以任意的精度近似任何一个连续型的密度.

其中 μ_l、σ_l、SK_l 以及 K_l 分别是分布 $p_l(x)$ 的均值、方差、斜度和峰度.

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 混合模型：正态 (高斯) 混合分布的概率密度函数为

$$f_{\mathrm{GM}}(x) = \sum_{l=1}^{L} \frac{w_l}{\sqrt{2\pi}\ \sigma_l}\ \mathrm{e}^{-\frac{(x-\mu_l)^2}{2\sigma_l^2}}$$

均值为 0 的正态分布构成的正态混合分布，其密度函数为

$$f_{\mathrm{GM}}(x) = \sum_{l=1}^{L} \frac{w_l}{\sqrt{2\pi}\ \sigma_l}\ \mathrm{e}^{-\frac{x^2}{2\sigma_l^2}}$$

均值 $\mu = 0$

方差 $\sigma^2 = \sum_{l=1}^{L} w_l \sigma_l^2$

斜度 $= 0$

峰度 $= \sum_{l=1}^{L} w_l \left(\frac{\sigma_l}{\sigma}\right)^4 \cdot 3$

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 多元广义双曲分布：特征函数

$$\Phi(\boldsymbol{t}) = \left(\frac{\alpha^2 - \boldsymbol{\beta}^{\mathrm{T}} \Delta \boldsymbol{\beta}}{\alpha^2 - \boldsymbol{\beta}^{\mathrm{T}} \Delta \boldsymbol{\beta} \ + \ \frac{1}{2} \boldsymbol{t}^{\mathrm{T}} \Delta \boldsymbol{t} - \mathbf{i} \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\Delta t}} \right)^{\frac{\lambda}{2}} \cdot \frac{K_\lambda \left(\delta \sqrt{\alpha^2 - \boldsymbol{\beta}^{\mathrm{T}} \Delta \boldsymbol{\beta} \ + \ \frac{1}{2} \boldsymbol{t}^{\mathrm{T}} \Delta \boldsymbol{t} - \mathbf{i} \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\Delta t}} \right)}{K_\lambda \left(\delta \sqrt{\alpha^2 - \boldsymbol{\beta}^{\mathrm{T}} \Delta \boldsymbol{\beta}} \right)}$$

多元双曲 (HYP) 分布：$\lambda = \dfrac{d+1}{2}$.

多元正态逆高斯 (NIG) 分布：$\lambda = -\dfrac{1}{2}$.

已学知识点 (Recap)

第 4 章 多元分布

4.6 厚尾分布

- 多元 t 分布：自由度为 n、非中心参数为 $\boldsymbol{\mu}$ 的非中心 t 分布的概率密度函数为

$$f_t(\boldsymbol{t};\ n,\ \Sigma,\ \boldsymbol{\mu}) = \frac{\Gamma\left(\frac{n+p}{2}\right)}{\Gamma\left(\frac{n}{2}\right) n^{\frac{p}{2}} \pi^{\frac{p}{2}} \left|\Sigma\right|^{\frac{1}{2}} \left[1 + \frac{1}{n}(\boldsymbol{t}-\boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{t}-\boldsymbol{\mu})\right]^{\frac{n+p}{2}}}$$

- 多元 Laplace 分布：概率密度函数为

$$K_\lambda = \frac{1}{2}\left(\frac{x}{2}\right)^{\lambda} \int_0^{\infty} t^{-\lambda-1} \mathrm{e}^{-t-\frac{x^2}{4t}} dt\,,\ x > 0$$

$$f_{M\mathrm{Laplace}_d}(\boldsymbol{x};\ \boldsymbol{m},\ \Sigma) = \frac{2\mathrm{e}^{\boldsymbol{x}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{m}}}{(2\pi)^{\frac{d}{2}}\left|\Sigma\right|^{\frac{1}{2}}} \left(\frac{\boldsymbol{x}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{x}}{2+\boldsymbol{m}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{m}}\right)^{\frac{\lambda}{2}} K_\lambda\left(\sqrt{\left(2+\boldsymbol{m}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{m}\right)\left(\boldsymbol{x}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{x}\right)}\right)$$

$$\lambda = \frac{2-d}{2}$$

$$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{m}$$
$$\mathrm{Var}(\boldsymbol{X}) = \Sigma + \boldsymbol{m}\boldsymbol{m}^{\mathrm{T}}$$

- 多元正态混合分布的概率密度函数为

$$f(\boldsymbol{x}) = \sum_{l=1}^{L} \frac{w_l}{\left|2\pi\Sigma_l\right|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_l)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_l)}$$

已学知识点 (Recap)

第 4 章 多元分布

4.8 自助法

- 样本容量较小时，自助法可以提升置信区间的精度.
- 自助分布 $\mathscr{L}\left\{\sqrt{n}\left(\frac{\overline{x}^{*}-\overline{x}}{\widehat{\sigma}^{*}}\right)\right\}$ 收敛于分布 $\mathscr{L}\left\{\sqrt{n}\left(\frac{\overline{x}^{*}-\overline{x}}{\widehat{\sigma}}\right)\right\}$ 的渐进极限分布.

Theory of the Multi-normal

多元正态分布的理论

概述

多元正态分布的理论

- 多元正态分布的基本性质
- Wishart 分布
- Hotelling's T^2 分布
- 球形分布和椭圆分布

多元正态分布的基本性质

- 上一章已建立的性质
  - $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ 的概率密度函数为

$$f(\boldsymbol{x}) = \left|2\pi\Sigma\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

$$\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu}$$

$$\mathbb{C}\mathrm{ov}(\boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right] = \Sigma$$

  - 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，矩阵 $\mathscr{A}_{p\times p}$，向量 $\boldsymbol{c} \in \mathbb{R}^p$，则

$$\boldsymbol{Y} = \mathscr{A}\boldsymbol{X} + \boldsymbol{c} \sim N_p\left(\mathscr{A}\boldsymbol{\mu} + \boldsymbol{c}, \mathscr{A}\Sigma\mathscr{A}^{\mathrm{T}}\right)$$

  - 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，经 Mahalanobis 变换则有

$$\boldsymbol{Y} = \Sigma^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}) \sim N_p\left(\boldsymbol{0}, \mathscr{I}_p\right),$$

    且有

$$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} = (\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \sim \chi_p^2$$

多元正态分布的基本性质

定理 5.1 设 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\boldsymbol{X}_1 \in \mathbb{R}^r$，$\boldsymbol{X}_2 \in \mathbb{R}^{p-r}$. 利用协方差矩阵的分块

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

定义

$$\boldsymbol{X}_{2\cdot 1} = \boldsymbol{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{X}_1$$

则

$$\boldsymbol{X}_1 \sim N_r(\boldsymbol{\mu}_1, \Sigma_{11})$$

$$\boldsymbol{X}_{2\cdot 1} \sim N_{p-r}(\boldsymbol{\mu}_{2\cdot 1}, \Sigma_{22\cdot 1})$$

相互独立，其中

$$\boldsymbol{\mu}_{2\cdot 1} = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1, \quad \Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

多元正态分布的基本性质

定理 5.1 设 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\boldsymbol{X}_1 \in \mathbb{R}^r$，$\boldsymbol{X}_2 \in \mathbb{R}^{p-r}$. 利用协方差矩阵的分块

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

定义

$$\boldsymbol{X}_{2\cdot 1} = \boldsymbol{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{X}_1$$

则

$$\boldsymbol{X}_1 \sim N_r(\boldsymbol{\mu}_1, \Sigma_{11})$$
$$\boldsymbol{X}_{2\cdot 1} \sim N_{p-r}(\boldsymbol{\mu}_{2\cdot 1}, \Sigma_{22\cdot 1})$$

相互独立，其中

$$\boldsymbol{\mu}_{2\cdot 1} = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1, \quad \Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

- 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，矩阵 $\mathscr{A}_{p\times p}$，向量 $\boldsymbol{c} \in \mathbb{R}^p$，则
$$\boldsymbol{Y} = \mathscr{A}\boldsymbol{X} + \boldsymbol{c} \sim N_p\left(\mathscr{A}\boldsymbol{\mu} + \boldsymbol{c}, \mathscr{A}\Sigma\mathscr{A}^{\mathrm{T}}\right)$$

证明： 记 $\mathscr{A} = (\mathscr{I}_r, \mathbf{0})$，$\mathscr{B} = \left(-\Sigma_{21}\Sigma_{11}^{-1}, \mathscr{I}_{p-r}\right)$，

$$\boldsymbol{X}_1 = (\mathscr{I}_r, \mathbf{0})\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} = \mathscr{A}\boldsymbol{X}$$

$$\begin{aligned} \boldsymbol{X}_{2\cdot 1} &= \boldsymbol{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{X}_1 \\ &= \left(-\Sigma_{21}\Sigma_{11}^{-1}, \mathscr{I}_{p-r}\right)\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \\ &= \mathscr{B}\boldsymbol{X} \end{aligned}$$

则 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_{2\cdot 1}$ 均服从正态分布.

$$\mathbb{C}\mathrm{ov}(\boldsymbol{X}_1, \boldsymbol{X}_{2\cdot 1}) = \mathscr{A}\Sigma\mathscr{B}^{\mathrm{T}} = \begin{pmatrix}\mathscr{I}_r & \mathbf{0}\end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} -\Sigma_{21}\Sigma_{11}^{-1} & \mathscr{I}_{p-r} \end{pmatrix}^{\mathrm{T}}$$

$$= \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\end{pmatrix}\begin{pmatrix} \left(-\Sigma_{21}\Sigma_{11}^{-1}\right)^{\mathrm{T}} \\ \mathscr{I}_{p-r} \end{pmatrix} = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\end{pmatrix}\begin{pmatrix} -\left(\Sigma_{11}^{-1}\right)^{\mathrm{T}}\Sigma_{21}^{\mathrm{T}} \\ \mathscr{I}_{p-r} \end{pmatrix} = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\end{pmatrix}\begin{pmatrix} -\Sigma_{11}^{-1}\Sigma_{12} \\ \mathscr{I}_{p-r} \end{pmatrix}$$

$$= -\Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12} + \Sigma_{12} \equiv \mathbf{0} \quad \Longrightarrow \quad \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_{2\cdot 1} \end{pmatrix} = \begin{pmatrix} \mathscr{A}\boldsymbol{X} \\ \mathscr{B}\boldsymbol{X} \end{pmatrix} = \begin{pmatrix} \mathscr{A} \\ \mathscr{B} \end{pmatrix}\boldsymbol{X} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_{2\cdot 1} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22\cdot 1} \end{pmatrix}\right)$$

多元正态分布的基本性质

证明：$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_{2\cdot 1} \end{pmatrix} = \begin{pmatrix} \mathscr{A} \\ \mathscr{B} \end{pmatrix} \boldsymbol{X} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_{2\cdot 1} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22\cdot 1} \end{pmatrix} \right)$

▸ $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ 的概率密度函数为

$$f(\boldsymbol{x}) = \left|2\pi\Sigma\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

$$f(\boldsymbol{x}_1, \boldsymbol{x}_{2\cdot 1}) = \left|2\pi \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22\cdot 1} \end{pmatrix}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left[\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_{2\cdot 1} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_{2\cdot 1} \end{pmatrix}\right]^{\mathrm{T}} \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22\cdot 1} \end{pmatrix}^{-1} \left[\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_{2\cdot 1} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_{2\cdot 1} \end{pmatrix}\right]\right\}$$

$$= \left|2\pi\Sigma_{11}\right|^{-1/2} \cdot \left|2\pi\Sigma_{22\cdot 1}\right|^{-1/2} \exp\left\{-\frac{1}{2}\begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Sigma_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22\cdot 1}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1} \end{pmatrix}\right\}$$

$$= \left|2\pi\Sigma_{11}\right|^{-1/2} \cdot \left|2\pi\Sigma_{22\cdot 1}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^{\mathrm{T}} \quad (\boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1})^{\mathrm{T}}\right] \begin{bmatrix} \Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \\ \Sigma_{22\cdot 1}^{-1}(\boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1}) \end{bmatrix}\right\}$$

$$= \left|2\pi\Sigma_{11}\right|^{-1/2} \cdot \left|2\pi\Sigma_{22\cdot 1}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^{\mathrm{T}} \Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) + (\boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1}) \Sigma_{22\cdot 1}^{-1}(\boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1})\right]\right\}$$

$$= \underbrace{\left|2\pi\Sigma_{11}\right|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^{\mathrm{T}} \Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)\right\}}_{f_{X_1}(\boldsymbol{x}_1)} \cdot$$

$$\underbrace{\left|2\pi\Sigma_{22\cdot 1}\right|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1}) \Sigma_{22\cdot 1}^{-1}(\boldsymbol{x}_{2\cdot 1} - \boldsymbol{\mu}_{2\cdot 1})\right\}}_{f_{X_{2\cdot 1}}(\boldsymbol{x}_{2\cdot 1})} = f_{X_1}(\boldsymbol{x}_1) \cdot f_{X_{2\cdot 1}}(\boldsymbol{x}_{2\cdot 1})$$

$\boldsymbol{X}_1$ 与 $\boldsymbol{X}_{2\cdot 1}$ 相互独立

多元正态分布的基本性质

推论 5.1 设 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$，则 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 相互独立的充分必要条件是 $\Sigma_{12} = \boldsymbol{0}$.

推论 5.2 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，对于给定的矩阵 $\mathscr{A}$ 与 $\mathscr{B}$，则 $\mathscr{A}\boldsymbol{X}$ 与 $\mathscr{B}\boldsymbol{X}$ 相互独立的充分必要条件是 $\mathscr{A}\Sigma\mathscr{B}^{\mathrm{T}} = \boldsymbol{0}$.

定理 5.2 设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，矩阵 $\mathscr{A}_{q\times p}$，向量 $\boldsymbol{c} \in \mathbb{R}^q$，且 $q \le p$，则 $\boldsymbol{Y} = \mathscr{A}\boldsymbol{X} + \boldsymbol{c}$ 服从 q 维正态分布，即

$$\boldsymbol{Y} \sim N_q(\mathscr{A}\boldsymbol{\mu} + \boldsymbol{c}, \mathscr{A}\Sigma\mathscr{A}^{\mathrm{T}}).$$

多元正态分布的基本性质

定理 5.3 给定 $X_1 = x_1$ 时 X_2 的条件分布亦是正态分布，其均值向量为

$\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$，协方差矩阵为 $\Sigma_{22\cdot 1}$，即

$$\left(X_2 \,\middle|\, X_1 = x_1\right) \sim N_{p-r}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)\ ,\ \Sigma_{22\cdot 1}\right)$$

证明：

$$X_{2\cdot 1} = X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1 \implies X_2 = X_{2\cdot 1} + \Sigma_{21}\Sigma_{11}^{-1}X_1$$

对于给定 $X_1 = x_1$ 的值，

$$\left(X_2 \,\middle|\, X_1 = x_1\right) = X_{2\cdot 1} + \Sigma_{21}\Sigma_{11}^{-1}x_1$$

$$X_{2\cdot 1} \sim N_{p-r}(\mu_{2\cdot 1}, \Sigma_{22\cdot 1}) \implies \left(X_2 \,\middle|\, X_1 = x_1\right) \sim N\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}x_1\ ,\ \Sigma_{22\cdot 1}\right)$$

X_1 的一个线性组合

与 X_1 无关

多元正态分布的基本性质

- 例: 设

$$p = 2,\ r = 1,\ \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \Sigma = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 2 \end{pmatrix}$$

则

定理 5.3 给定 $X_1 = x_1$ 时 X_2 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)$，协方差矩阵为 $\Sigma_{22\cdot 1}$，即

$$(X_2 | X_1 = x_1) \sim N_{p-r}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1),\ \Sigma_{22\cdot 1}\right)$$

$$\Sigma_{11} = 1\ ,\quad \Sigma_{21} = -0.8\ ,$$

$$\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 2 - (-0.8) \times 1 \times (-0.8) = 1.36$$

X_1 的边际概率密度函数为

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right)$$

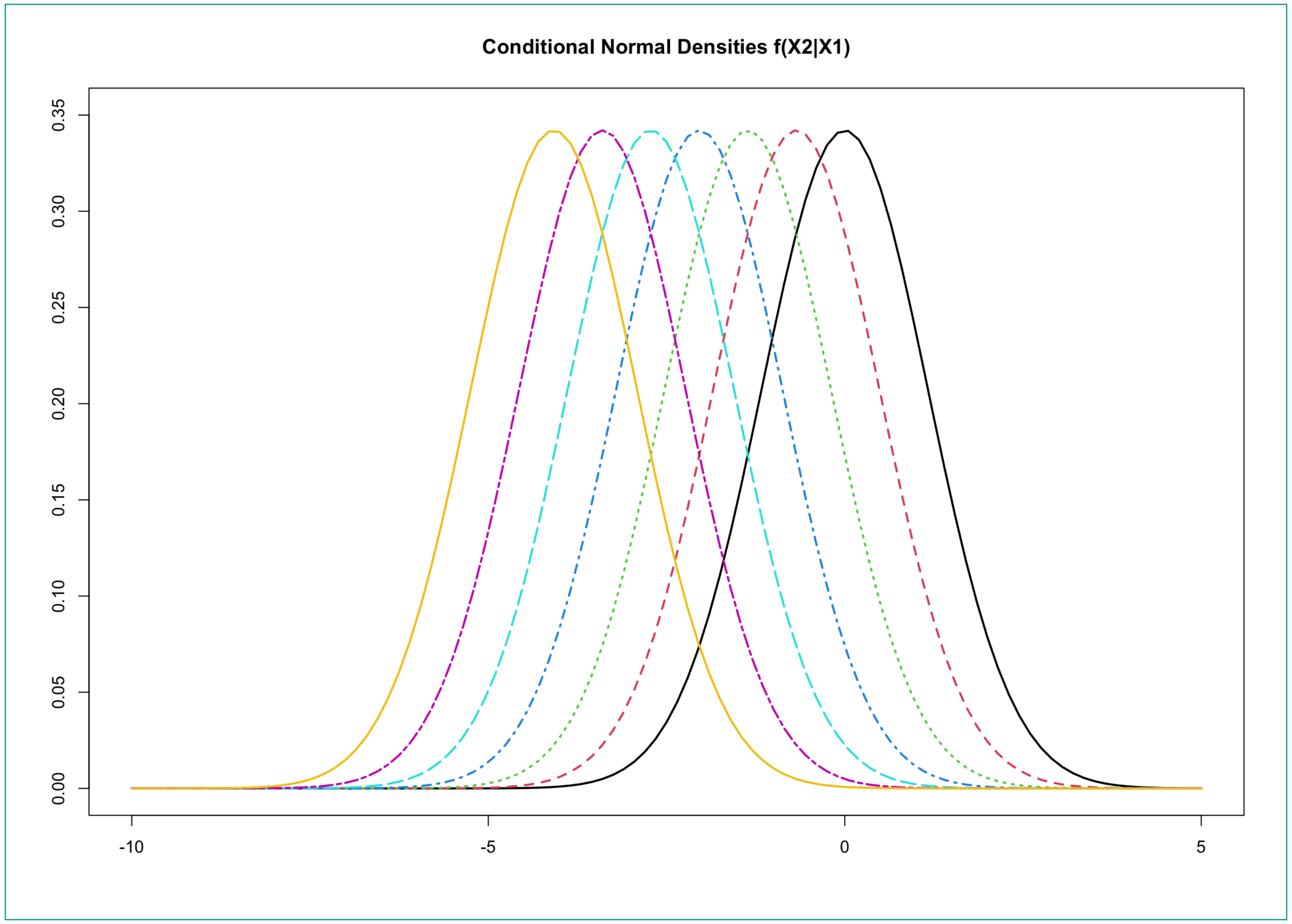
$(X_2 | X_1 = x_1)$ 的条件概率密度函数为

条件均值 $\mathrm{E}(X_2 | X_1 = x_1) = -0.8x_1$ 是 X_1 的线性函数.

$$f(x_2 | x_1) = \frac{1}{\sqrt{2\pi \times (1.36)}} \exp\left(-\frac{(x_2 + 0.8x_1)^2}{2 \times (1.36)}\right)$$

多元正态分布的基本性质

```
x = 0
mu = -0.8 * x
sigma = sqrt(1.36)
curve(dnorm(x, mu, sigma), -10, 5, col=1, xlab='', ylab='', lwd=2, ylim=c(0, 0.35),
     main='Conditional Normal Densities f(X2|X1)')
for (i in 1:6) {
  x = x + 0.85
  mu = -0.8 * x
  curve(dnorm(x, mu, sigma), -10, 5, col=i+1, xlab='', ylab='', add=TRUE, lty=i+1, lwd=2)
}
```



多元正态分布的基本性质

定理 5.4 设 $\boldsymbol{X}_1 \sim N_r(\boldsymbol{\mu}_1, \Sigma_{11})$，$\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}(\mathscr{A}\boldsymbol{x}_1 + \boldsymbol{b}, \Omega)$，其中 Ω 不依赖于 $\boldsymbol{x}_1$，则

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p(\boldsymbol{\mu}, \Sigma)$$

其中

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mathscr{A}\boldsymbol{\mu}_1 + \boldsymbol{b} \end{pmatrix},\quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11}\mathscr{A}^{\mathrm{T}} \\ \mathscr{A}\Sigma_{11} & \Omega + \Sigma_{11}\mathscr{A}^{\mathrm{T}} \end{pmatrix}$$

多元正态分布的基本性质

- **例**：考虑下述随机变量

$$X_1 \sim N_1(0,\ 1)\ ,\quad \left(\boldsymbol{X}_2 \,\middle|\, X_1 = x_1\right) \sim N_2\left(\begin{pmatrix} 2x_1 \\ x_1 + 1 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

$$\mathscr{A} = \begin{pmatrix} 2 \\ 1 \end{pmatrix},\quad \boldsymbol{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\quad \Omega = \mathscr{I}_2$$

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},\ \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{pmatrix}\right)$$

定理 5.4 设 $\boldsymbol{X}_1 \sim N_r(\boldsymbol{\mu}_1,\ \Sigma_{11})$，$\left(\boldsymbol{X}_2 \,\middle|\, \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}(\mathscr{A}\boldsymbol{x}_1 + \boldsymbol{b},\ \Omega)$，其中 Ω 不依赖于 $\boldsymbol{x}_1$，则

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p(\boldsymbol{\mu},\ \Sigma)$$

其中

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mathscr{A}\boldsymbol{\mu}_1 + \boldsymbol{b} \end{pmatrix},\quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11}\mathscr{A}^{\mathrm{T}} \\ \mathscr{A}\Sigma_{11} & \Omega + \Sigma_{11}\mathscr{A}^{\mathrm{T}} \end{pmatrix}$$

$$\Longrightarrow \quad \boldsymbol{X}_2 \sim N_2\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\right)$$

- 在 X_1 给定的条件下，$\boldsymbol{X}_2$ 的两个分量相互独立.
- 但从 $\boldsymbol{X}_2$ 的边际分布来看，它的两个分量不相互独立.

多元正态分布的基本性质

- 例：考虑下述随机变量

$$X_1 \sim N_1(0,\ 1)\ ,\quad \left(\boldsymbol{X}_2 \,\middle|\, X_1 = x_1\right) \sim N_2\left(\begin{pmatrix} 2x_1 \\ x_1 + 1 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

- $\boldsymbol{X}_2$ 的边际分布的均值向量与协方差矩阵可计算如下

$$\mathrm{E}\left(\boldsymbol{X}_2\right) = \mathrm{E}\left[\mathrm{E}\left(\boldsymbol{X}_2 \,\middle|\, X_1\right)\right] = \mathrm{E}\left[\begin{pmatrix} 2X_1 \\ X_1 + 1 \end{pmatrix}\right] = \begin{pmatrix} \mathrm{E}\left(2X_1\right) \\ \mathrm{E}\left(X_1\right) + 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{aligned} \mathrm{Var}\left(\boldsymbol{X}_2\right) &= \mathrm{E}\left[\mathrm{Var}\left(\boldsymbol{X}_2 \,\middle|\, X_1\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\boldsymbol{X}_2 \,\middle|\, X_1\right)\right] \\ &= \mathrm{E}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] + \mathrm{Var}\left[\begin{pmatrix} 2X_1 \\ X_1 + 1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} \end{aligned}$$

多元正态分布的基本性质

- 条件逼近

定理 4.3 设 $\boldsymbol{X}_1 \in \mathbb{R}^k$，$\boldsymbol{X}_2 \in \mathbb{R}^{p-k}$，$\boldsymbol{U} = \boldsymbol{X}_2 - \mathrm{E}\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1\right)$. 则

1. $\mathrm{E}(\boldsymbol{U}) = \mathbf{0}$.
2. 用 $\boldsymbol{X}_1$ 的一个函数 $h\left(\boldsymbol{X}_1\right)$ 对 $\boldsymbol{X}_2$ 作近似，则 $\mathrm{E}\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1\right)$ 是一个“**最佳**”逼近，其中 h： $\mathbb{R}^k \to \mathbb{R}^{p-k}$. 所谓“**最佳**”是指**均方误差** (MSE) 最小的意义下，其中

$$\mathrm{MSE}(h) = \mathrm{E}\left\{\left[\boldsymbol{X}_2 - h\left(\boldsymbol{X}_1\right)\right]^{\mathrm{T}}\left[\boldsymbol{X}_2 - h\left(\boldsymbol{X}_1\right)\right]\right\}.$$

定理 5.3 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1\right)$，协方差矩阵为 $\Sigma_{22\cdot1}$，即

$$\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{\mu}_1\right)\ , \Sigma_{22\cdot1}\right)$$

$$\begin{aligned}
\Longrightarrow \quad \boldsymbol{X}_2 &= \mathrm{E}\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1\right) + \boldsymbol{U} \\
&= \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\boldsymbol{X}_1 - \boldsymbol{\mu}_1\right) + \boldsymbol{U} \\
&= \underbrace{\boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1}_{\boldsymbol{\beta}_0} + \underbrace{\Sigma_{21}\Sigma_{11}^{-1}}_{\mathscr{B}}\boldsymbol{X}_1 + \boldsymbol{U} \\
&= \boldsymbol{\beta}_0 + \mathscr{B}\boldsymbol{X}_1 + \underbrace{\boldsymbol{U}}_{N\left(\mathbf{0},\ \Sigma_{22\cdot1}\right)}
\end{aligned}$$

多元正态分布的基本性质

- 条件逼近

$$\mathscr{B} = \Sigma_{21}\Sigma_{11}^{-1}$$

$$\boldsymbol{X}_2 = \boldsymbol{\beta}_0 + \mathscr{B}\boldsymbol{X}_1 + \boldsymbol{U}$$

$$\boldsymbol{\beta}_0 = \boldsymbol{\mu}_2 - \mathscr{B}\boldsymbol{\mu}_1 \qquad N\left(\boldsymbol{0},\ \Sigma_{22\cdot 1}\right)$$

  - 一种特殊情形： $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ X_2 \end{pmatrix}_{p\times 1}$ ， $X_2 \in \mathbb{R}$

$$\boldsymbol{\beta}^{\mathrm{T}} = \mathscr{B} = \left(\boldsymbol{\sigma}_{21}\right)_{1\times(p-1)}\left(\Sigma_{11}^{-1}\right)_{(p-1)\times(p-1)}$$

$$\Longrightarrow \quad X_2 = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_1 + U$$

$$\beta_0 = \mu_2 - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\mu}_1 \qquad N\left(0,\ \sigma_{22\cdot 1}\right)$$

  - 用 $\boldsymbol{X}_1$ 的一个函数对 X_2 的最佳 MSE 逼近是一个超平面.
  - X_2 的边际分布的方差为

$$\sigma_{22} = \boldsymbol{\beta}^{\mathrm{T}}\Sigma_{11}\boldsymbol{\beta} + \sigma_{22\cdot 1} = \boldsymbol{\sigma}_{21}\Sigma_{11}^{-1}\Sigma_{11}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{12} + \sigma_{22\cdot 1} = \boldsymbol{\sigma}_{21}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{12} + \sigma_{22\cdot 1}$$

$$\Longrightarrow \quad \frac{\boldsymbol{\sigma}_{21}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{12}}{\sigma_{22}} + \frac{\sigma_{22\cdot 1}}{\sigma_{22}} = 1$$

多元正态分布的基本性质

- 条件逼近

$$\boxed{\frac{\boldsymbol{\sigma}_{21}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{12}}{\sigma_{22}}} + \frac{\sigma_{22\cdot 1}}{\sigma_{22}} = 1$$

- $\rho^2_{2\cdot 1\ldots(p-1)}$：$X_2$ 与 r 个变量 $\boldsymbol{X}_1$ 之间的多重相关系数的平方.
- $\rho^2_{2\cdot 1\ldots(p-1)}$：$X_2$ 的方差可以由线性逼近 $\beta_0+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_1$ 进行解释的比例.
- $\rho^2_{2\cdot 1\ldots(p-1)}$：$X_2$ 与 $\boldsymbol{X}_1$ 分量的线性组合之间的相关系数能够达到的最大值.
- $\rho^2_{2\cdot 1\ldots(p-1)}$：当 $p=2$ 时，多重相关系数 ρ_{21} 和 X_2 与 X_1 常规的简单相关系数恰好相同.

多元正态分布的基本性质

- **例**：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\mu = \begin{pmatrix} 172.7 \\ 104.6 \\ 104.0 \\ 93.8 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1037.21 & -80.02 & 1430.7 & 271.44 \\ -80.02 & 219.84 & 92.1 & -91.58 \\ 1430.70 & 92.10 & 2624.0 & 210.30 \\ 271.44 & -91.58 & 210.3 & 177.36 \end{pmatrix}$$

```
> pullover
    X1  X2  X3  X4
1  230 125 200 109
2  181  99  55 107
3  165  97 105  98
4  150 115  85  71
5   97 120   0  82
6  192 100 150 103
7  181  80  85 111
8  189  90 120  93
9  172  95 110  86
10 170 125 130  78
```

```
x <- c(230, 125, 200, 109, 181, 99, 55, 107, 165, 97, 105, 98, 150, 115, 85, 71, 97, 120, 0, 82,
    192, 100, 150, 103, 181, 80, 85, 111, 189, 90, 120, 93, 172, 95, 110, 86, 170, 125, 130, 78)
pullover <- matrix(x, ncol = 4, byrow = TRUE)
pullover <- as.data.frame(pullover)
colnames(pullover) <- c("X1", "X2", "X3", "X4")
pullover
mu = apply(pullover, 2, mean)
mu
Sigma = var(pullover) * (9 / 10)
Sigma
```

```
> apply(pullover, 2, mean)
   X1    X2    X3    X4
172.7 104.6 104.0  93.8
```

```
> var(pullover) * (9 / 10)
        X1     X2     X3     X4
X1 1037.21 -80.02 1430.7 271.44
X2  -80.02 219.84   92.1 -91.58
X3 1430.70  92.10 2624.0 210.30
X4  271.44 -91.58  210.3 177.36
```

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$
\boldsymbol{\mu} = \begin{array}{c} \\ \begin{pmatrix} 172.7 \\ 104.6 \\ 104.0 \\ 93.8 \end{pmatrix} \end{array}
\begin{array}{l} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{array}, \quad
\Sigma = \begin{array}{cc} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \multicolumn{2}{c}{\begin{pmatrix} 1037.21 & -80.02 & 1430.7 & 271.44 \\ -80.02 & 219.84 & 92.1 & -91.58 \\ 1430.70 & 92.10 & 2624.0 & 210.30 \\ 271.44 & -91.58 & 210.3 & 177.36 \end{pmatrix}} \\ \boldsymbol{\sigma}_{21} & \Sigma_{22} \end{array}
$$

  - 则给定 (X_2, X_3, X_4) 时 X_1 的条件分布服从一元正态分布，其均值为

$$
\mu_1 + \boldsymbol{\sigma}_{12}\Sigma_{22}^{-1}\begin{pmatrix} X_2 - \mu_2 \\ X_3 - \mu_3 \\ X_4 - \mu_4 \end{pmatrix}
$$

定理 5.3 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)$，协方差矩阵为 $\Sigma_{22\cdot 1}$，即

$$
\left(\boldsymbol{X}_2 \middle| \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1),\ \Sigma_{22\cdot 1}\right)
$$

$$
= 172.7 + (-80.02 \quad 1430.70 \quad 271.44)\begin{pmatrix} 219.84 & 92.10 & -91.58 \\ 92.10 & 2624.00 & 210.30 \\ -91.58 & 210.30 & 177.36 \end{pmatrix}^{-1}\begin{pmatrix} X_2 - 104.6 \\ X_3 - 104.0 \\ X_4 - 93.8 \end{pmatrix}
$$

$$
= 65.670 - 0.216X_2 + 0.485X_3 + 0.844X_4
$$

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\boldsymbol{\mu}=\begin{pmatrix}172.7\\104.6\\104.0\\93.8\end{pmatrix}\begin{array}{l}\mu_1\\\mu_2\\\mu_3\\\mu_4\end{array},\quad \Sigma=\begin{array}{c}\begin{array}{cc}\sigma_{11} & \sigma_{12}\end{array}\\ \left(\begin{array}{c|cccc}1037.21 & -80.02 & 1430.7 & 271.44\\\hline -80.02 & 219.84 & 92.1 & -91.58\\1430.70 & 92.10 & 2624.0 & 210.30\\271.44 & -91.58 & 210.3 & 177.36\end{array}\right)\\ \begin{array}{cc}\sigma_{21} & \Sigma_{22}\end{array}\end{array}$$

  - 则给定 (X_2, X_3, X_4) 时 X_1 的条件分布服从一元正态分布，其均值为

```
mu_2 = as.matrix(mu[2:4])
Sigma_21 = as.matrix(Sigma[1, 2:4])
Sigma_22 = as.matrix(Sigma[2:4, 2:4])
t(Sigma_21) %*% solve(Sigma_22)
mu[1] - t(Sigma_21) %*% solve(Sigma_22) %*% mu_2
```

```
> t(Sigma_21) %*% solve(Sigma_22)
             X2        X3        X4
[1,] -0.2157821 0.4851898 0.8437261
> mu[1] - t(Sigma_21) %*% solve(Sigma_22) %*% mu_2
         [,1]
[1,] 65.66956
```

$$= 65.670 - 0.216X_2 + 0.485X_3 + 0.844X_4$$

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\boldsymbol{\mu}=\begin{array}{c}\begin{pmatrix}172.7\\ \hdashline 104.6\\104.0\\93.8\end{pmatrix}\end{array}\begin{array}{l}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{array}\qquad \Sigma=\begin{array}{cc}\sigma_{11} & \boldsymbol{\sigma}_{12}\\ \multicolumn{2}{c}{\left(\begin{array}{c:ccc}1037.21 & -80.02 & 1430.7 & 271.44\\ \hdashline -80.02 & 219.84 & 92.1 & -91.58\\ 1430.70 & 92.10 & 2624.0 & 210.30\\ 271.44 & -91.58 & 210.3 & 177.36\end{array}\right)}\\ \boldsymbol{\sigma}_{21} & \Sigma_{22}\end{array}$$

- 则给定 (X_2, X_3, X_4) 时 X_1 的条件分布服从一元正态分布，其方差为

$$\sigma_{11\cdot 2}=\sigma_{11}-\boldsymbol{\sigma}_{12}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$$

定理 5.3 给定 $\boldsymbol{X}_1=\boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2+\Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)$，协方差矩阵为 $\Sigma_{22\cdot1}$，即

$$\left(\boldsymbol{X}_2\,|\,\boldsymbol{X}_1=\boldsymbol{x}_1\right)\sim N_{p-r}\left(\boldsymbol{\mu}_2+\Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)\ ,\Sigma_{22\cdot1}\right)$$

$$=1037.21-(-80.02\quad 1430.70\quad 271.44)\begin{pmatrix}219.84 & 92.10 & -91.58\\ 92.10 & 2624.00 & 210.30\\ -91.58 & 210.30 & 177.36\end{pmatrix}^{-1}\begin{pmatrix}-80.02\\1430.70\\271.44\end{pmatrix}=96.761$$

```
Sigma[1, 1] - t(Sigma_21) %*% solve(Sigma_22) %*% Sigma_12
```

```
> Sigma[1, 1] - t(Sigma_21) %*% solve(Sigma_22) %*% Sigma_21
         [,1]
[1,] 96.76111
```

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\boldsymbol{\mu}=\begin{pmatrix}172.7\\104.6\\104.0\\93.8\end{pmatrix}\begin{matrix}\mu_1\\\mu_2\\\mu_3\\\mu_4\end{matrix};\quad \Sigma=\begin{pmatrix}1037.21 & -80.02 & 1430.7 & 271.44\\-80.02 & 219.84 & 92.1 & -91.58\\1430.70 & 92.10 & 2624.0 & 210.30\\271.44 & -91.58 & 210.3 & 177.36\end{pmatrix}$$

σ_{11}, $\boldsymbol{\sigma}_{12}$, $\boldsymbol{\sigma}_{21}$, Σ_{22}

  - 价格 (X_2)、广告费用 (X_3) 和促销员费用 (X_4) 对销售量 (X_1) 的线性逼近由条件均值给出

$$65.670-0.216X_2+0.485X_3+0.844X_4$$

  - 逼近的效果由多重相关系数给出

$$\rho^2_{1\cdot 234}=\frac{\boldsymbol{\sigma}_{12}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}}{\sigma_{11}}=0.907$$

```
t(Sigma_21) %*% solve(Sigma_22) %*% Sigma_21 / Sigma[1, 1]
```

```
> t(Sigma_21) %*% solve(Sigma_22) %*% Sigma_21 / Sigma[1, 1]
          [,1]
[1,] 0.9067102
```

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\boldsymbol{\mu} = \begin{pmatrix} 172.7 \\ 104.6 \\ \hline 104.0 \\ 93.8 \end{pmatrix} \begin{matrix} \boldsymbol{\mu}_1 \\ \\ \boldsymbol{\mu}_2 \end{matrix}, \quad \Sigma = \left(\begin{array}{cc|cc} 1037.21 & -80.02 & 1430.7 & 271.44 \\ -80.02 & 219.84 & 92.1 & -91.58 \\ \hline 1430.70 & 92.10 & 2624.0 & 210.30 \\ 271.44 & -91.58 & 210.3 & 177.36 \end{array}\right) \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix}$$

  - 给定 (X_3, X_4) 时，(X_1, X_2) 的条件分布为二元正态分布，均值向量为

$$\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$$

定理 5.3 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)$，协方差矩阵为 $\Sigma_{22\cdot 1}$，即

$$\left(\boldsymbol{X}_2 \,\middle|\, \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1),\ \Sigma_{22\cdot 1}\right)$$

$$= \begin{pmatrix} 172.7 \\ 104.6 \end{pmatrix} + \begin{pmatrix} 1430.70 & 271.44 \\ 92.10 & -91.58 \end{pmatrix} \begin{pmatrix} 2624.00 & 210.30 \\ 210.30 & 177.36 \end{pmatrix}^{-1} \begin{pmatrix} X_3 - 104.0 \\ X_4 - 93.8 \end{pmatrix} = \begin{pmatrix} 32.516 + 0.467X_3 + 0.977X_4 \\ 153.644 + 0.085X_3 - 0.617X_4 \end{pmatrix}$$

```
mu_1 = as.matrix(mu[1:2])
mu_2 = as.matrix(mu[3:4])
Sigma_11 = as.matrix(Sigma[1:2, 1:2])
Sigma_12 = as.matrix(Sigma[1:2, 3:4])
Sigma_21 = as.matrix(Sigma[3:4, 1:2])
Sigma_22 = as.matrix(Sigma[3:4, 3:4])
Sigma_12 %*% solve(Sigma_22)
mu_1 - Sigma_12 %*% solve(Sigma_22) %*% mu_2
```

```
> Sigma_12 %*% solve(Sigma_22)
           X3         X4
X1 0.46695332  0.9767688
X2 0.08451319 -0.6165602
> mu_1 - Sigma_12 %*% solve(Sigma_22) %*% mu_2
        [,1]
X1  32.51594
X2 153.64398
```

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\boldsymbol{\mu} = \left(\begin{array}{c} 172.7 \\ 104.6 \\ \hline 104.0 \\ 93.8 \end{array}\right) \begin{array}{l} \boldsymbol{\mu}_1 \\ \\ \boldsymbol{\mu}_2 \end{array}, \quad \Sigma = \begin{array}{c} \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \end{array} \\ \left(\begin{array}{cc|cc} 1037.21 & -80.02 & 1430.7 & 271.44 \\ -80.02 & 219.84 & 92.1 & -91.58 \\ \hline 1430.70 & 92.10 & 2624.0 & 210.30 \\ 271.44 & -91.58 & 210.3 & 177.36 \end{array}\right) \\ \begin{array}{cc} \Sigma_{21} & \Sigma_{22} \end{array} \end{array}$$

  - 给定 (X_3, X_4) 时，(X_1, X_2) 的条件分布为二元正态分布，协方差矩阵为

$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

定理 5.3 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)$，协方差矩阵为 $\Sigma_{22\cdot 1}$，即

$$\left(\boldsymbol{X}_2 \,\middle|\, \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)\ , \Sigma_{22\cdot 1}\right)$$

$$= \begin{pmatrix} 1037.21 & -80.02 \\ -80.02 & 219.84 \end{pmatrix} - \begin{pmatrix} 1430.70 & 271.44 \\ 92.10 & -91.58 \end{pmatrix} \begin{pmatrix} 2624.00 & 210.30 \\ 210.30 & 177.36 \end{pmatrix}^{-1} \begin{pmatrix} 1430.70 & 92.10 \\ 271.44 & -91.58 \end{pmatrix}$$

$$= \begin{pmatrix} 104.006 & -33.574 \\ -33.574 & 155.592 \end{pmatrix}$$

```
> Sigma_11 - Sigma_12 %*% solve(Sigma_22) %*% Sigma_21
          X1        X2
X1 104.00576 -33.57391
X2 -33.57391 155.59175
```

```
Sigma_11 - Sigma_12 %*% solve(Sigma_22) %*% Sigma_21
```

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集，设 X_1 (销量)、X_2 (价格)、X_3 (广告费用) 和 X_4 (促销员费用) 服从正态分布，其中

$$\mu = \begin{pmatrix} 172.7 \\ 104.6 \\ 104.0 \\ 93.8 \end{pmatrix} \begin{matrix} \mu_1 \\ \\ \mu_2 \end{matrix}, \quad \Sigma = \left(\begin{array}{cc|cc} 1037.21 & -80.02 & 1430.7 & 271.44 \\ -80.02 & 219.84 & 92.1 & -91.58 \\ \hline 1430.70 & 92.10 & 2624.0 & 210.30 \\ 271.44 & -91.58 & 210.3 & 177.36 \end{array}\right) \quad \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix}$$

  - 给定 (X_3, X_4) 时，(X_1, X_2) 的条件分布为二元正态分布，协方差矩阵为

$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

定理 5.3 给定 $\boldsymbol{X}_1 = \boldsymbol{x}_1$ 时 $\boldsymbol{X}_2$ 的条件分布亦是正态分布，其均值向量为 $\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)$，协方差矩阵为 $\Sigma_{22\cdot 1}$，即

$$\left(\boldsymbol{X}_2 \,\middle|\, \boldsymbol{X}_1 = \boldsymbol{x}_1\right) \sim N_{p-r}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)\ ,\ \Sigma_{22\cdot 1}\right)$$

$$= \begin{pmatrix} 1037.21 & -80.02 \\ -80.02 & 219.84 \end{pmatrix} - \begin{pmatrix} 1430.70 & 271.44 \\ 92.10 & -91.58 \end{pmatrix} \begin{pmatrix} 2624.00 & 210.30 \\ 210.30 & 177.36 \end{pmatrix}^{-1} \begin{pmatrix} 1430.70 & 92.10 \\ 271.44 & -91.58 \end{pmatrix}$$

$$= \begin{pmatrix} 104.006 & -33.574 \\ -33.574 & 155.592 \end{pmatrix} \Longrightarrow \rho_{X_1X_2|X_3X_4} = \frac{-33.574}{\sqrt{104.006 \times 155.592}} = -0.264$$

给定 X_3 与 X_4 的取值时，X_1 与 X_2 的偏相关系数.

多元正态分布的基本性质

- 例：考虑“经典蓝”套头衫数据集.
  - 4 个变量的相关矩阵为

$$P = \begin{pmatrix} 1 & -0.168 & 0.867 & 0.633 \\ -0.168 & 1 & 0.121 & -0.464 \\ 0.867 & 0.121 & 1 & 0.308 \\ 0.633 & -0.464 & 0.308 & 1 \end{pmatrix}$$

```
> cor(pullover)
           X1         X2        X3         X4
X1  1.0000000 -0.1675760 0.8672280  0.6328673
X2 -0.1675760  1.0000000 0.1212619 -0.4637879
X3  0.8672280  0.1212619 1.0000000  0.3082688
X4  0.6328673 -0.4637879 0.3082688  1.0000000
```

```
cor(pullover)
```

$$\Longrightarrow \quad \rho_{X_1X_2} = -0.168$$

$$\Longrightarrow \quad \rho_{X_1X_2|X_3X_4} = -0.264$$

Wishart 分布

$$\text{数据矩阵: } \mathcal{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{pmatrix}.$$

$$\text{均值向量: } \bar{\boldsymbol{x}} = n^{-1}\mathcal{X}^{\mathrm{T}}\mathbf{1}_n$$

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1}$$

$$\begin{aligned} \text{样本协方差矩阵: } \mathcal{S} &= \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{x}_k - \bar{\boldsymbol{x}}\right)\left(\boldsymbol{x}_k - \bar{\boldsymbol{x}}\right)^{\mathrm{T}} \\ &= \frac{1}{n}\mathcal{X}^{\mathrm{T}}\mathcal{X} - \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathrm{T}} \\ &= \frac{1}{n}\mathcal{X}^{\mathrm{T}}\mathcal{H}\mathcal{X} \end{aligned}$$

$$\text{中心化矩阵: } \mathcal{H} = \mathcal{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$$

- Wishart 分布在估计协方差矩阵的分析中发挥着重要作用.

Wishart 分布

定义 5.1 随机矩阵 $\mathscr{X} = \left(X_{ij}\right)_{m\times n}$ 的分布是指其所有元素构成的随机向量

$$\boldsymbol{X} = \left(X_{11}, \ldots, X_{1n}, X_{21}, \ldots, X_{2n}, \ldots, X_{m1}, \ldots, X_{mn}\right)$$

的联合分布.

- 当 $m = n$ 且 $\mathscr{X} = \left(X_{ij}\right)_{n\times n}$ 为对称矩阵时，我们只需考虑其下三角矩阵的元素

$$\mathscr{X} = \begin{pmatrix} X_{11} & & & \\ X_{21} & X_{22} & & \\ \vdots & \vdots & \ddots & \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{pmatrix}_{n\times n}$$

即，$\boldsymbol{X} = \left(X_{11}, X_{21}, \ldots, X_{n1}, X_{22}, \ldots, X_{n2}, \ldots, X_{nn}\right)$ 的分布.

Wishart 分布

- 数据矩阵 $\mathscr{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{pmatrix}.$

- 假设 $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, 2, \ldots, n$

$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{pmatrix}$, $i = 1, 2, \ldots, n$

- 记 $\mathscr{M} = \mathbb{E}(\mathscr{X}) = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1^{\mathrm{T}} \\ \boldsymbol{\mu}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\mu}_n^{\mathrm{T}} \end{pmatrix}$

定义 5.2 设 $\mathscr{A} = \mathscr{X}^{\mathrm{T}}\mathscr{X} = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}$,

则称 $\mathscr{A}$ 服从非中心 Wishart 分布

$W_p(\Sigma, n, \boldsymbol{\tau})$，其中 $\boldsymbol{\tau} = \mathscr{M}^{\mathrm{T}}\mathscr{M}$ 是非

中心参数.

- 当 $\boldsymbol{\mu}_i = \mathbf{0}$, $i = 1, 2, \ldots, n$ 时，有 $\boldsymbol{\tau} = \mathbf{0}$，则称 $\mathscr{A}$ 服从中心 Wishart 分布，记作

$\mathscr{A} \sim W_p(\Sigma, n)$.

Wishart 分布

- 例：取 $p=1$，则对 $X \sim N_1\left(0, \sigma^2\right)$，数据矩阵为

$$\mathscr{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \Longrightarrow \mathscr{A} = \mathscr{X}^{\mathrm{T}}\mathscr{X} = (x_1 \quad x_2 \quad \cdots \quad x_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} x_i^2$$

$$= \sigma^2 \sum_{i=1}^{n} \left[\left(\frac{x_i}{\sigma}\right)^2\right]$$

$$= \sigma^2 \cdot \chi_n^2$$

Wishart 分布

定理 5.5 如果 $\mathscr{A} \sim W_p(\Sigma, n)$，矩阵 $\mathscr{B}_{p\times q}$，则 $\mathscr{B}^{\mathrm{T}}\mathscr{A}\mathscr{B}$ 服从 Wishart 分布$W_q\left(\mathscr{B}^{\mathrm{T}}\Sigma\mathscr{B}, n\right)$.

$\mathscr{A}$ 的线性变换

- 若取 $\mathscr{B} = \Sigma^{-1/2}$，则 $\Sigma^{-1/2}\mathscr{A}\Sigma^{-1/2}$ 即服从标准 Wishart 分布 $W_p(\mathscr{I}, n)$.

定理 5.6 设 $\mathscr{A} \sim W_p(\Sigma, m)$，$\boldsymbol{a} \in \mathbb{R}^p$ 满足 $\boldsymbol{a}^{\mathrm{T}}\Sigma\,\boldsymbol{a} \neq 0$，则 $\dfrac{\boldsymbol{a}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\Sigma\,\boldsymbol{a}}$ 服从 χ_m^2 分布.

$$\boldsymbol{x}_i \overset{\text{i.i.d.}}{\sim} N_p\left(\mathbf{0}, \Sigma\right) \implies \mathscr{A} = \mathscr{X}^{\mathrm{T}}\mathscr{X} = \sum_{i=1}^{m} \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} \sim W_p(\Sigma, m)$$

$$\implies \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_i \sim N_1\left(\mathbf{0}, \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}\right)$$

$$\implies \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_i}{\left(\boldsymbol{a}^{\mathrm{T}}\Sigma\,\boldsymbol{a}\right)^{1/2}} \sim N_1\left(\mathbf{0}, 1\right) \implies \frac{\boldsymbol{a}^{\mathrm{T}}\mathscr{A}\,\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\Sigma\,\boldsymbol{a}} = \sum_{i=1}^{m}\left[\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_i}{\left(\boldsymbol{a}^{\mathrm{T}}\Sigma\,\boldsymbol{a}\right)^{1/2}} \cdot \frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{a}}{\left(\boldsymbol{a}^{\mathrm{T}}\Sigma\,\boldsymbol{a}\right)^{1/2}}\right] \sim \chi_m^2$$

Wishart 分布

定理 5.7 (Cochran) 设 $\mathscr{X}_{n\times p}$ 是来自正态分布 $N_p\left(\mathbf{0}, \Sigma\right)$ 的一个数据矩阵，设 $\mathscr{C}_{n\times n}$ 是一个对称矩阵. 则

(a) $\mathscr{X}^{\mathrm{T}}\mathscr{C}\mathscr{X}$ 服从加权 Wishart 随机变量和的分布，即

$$\mathscr{X}^{\mathrm{T}}\mathscr{C}\mathscr{X} = \sum_{i=1}^{n} \lambda_i W_p(\Sigma, 1)$$

其中 $\lambda_i\ (i = 1, 2, \ldots, n)$ 是 $\mathscr{C}$ 的特征值.

(b) $\mathscr{X}^{\mathrm{T}}\mathscr{C}\mathscr{X}$ 服从 Wishart 分布的充分必要条件是 $\mathscr{C}^2 = \mathscr{C}$. 此时有

$$\mathscr{X}^{\mathrm{T}}\mathscr{C}\mathscr{X} \sim W_p(\Sigma, r)$$

其中 $r = \mathrm{rank}(\mathscr{C}) = \mathrm{tr}(\mathscr{C})$.

(c) $n\mathscr{S} = \mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$ 服从 Wishart 分布 $W_p(\Sigma, n-1)$.

$\mathscr{H} = \mathscr{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$

(d) $\bar{\boldsymbol{x}}$ 与 $\mathscr{S}$ 相互独立.

$\mathscr{S} = \frac{1}{n}\mathscr{X}^{\mathrm{T}}\mathscr{H}\mathscr{X}$